PIONEER: PANORAMIC IMAGINED OCCUPANCY-GUIDED ACTIVE MAPPING FOR EFFICIENT AERIAL EXPLORATION

Anonymous authors
Paper under double-blind review



Figure 1: **Overview of PIONEER.** Near-complete panoramic coverage enables position-centric planning. The proposed method predicts unseen occupancy to estimate trajectory coverage gains and implements dynamic beam search to achieve more efficient aerial exploration.

ABSTRACT

Active mapping seeks informative observations for efficient aerial exploration and reconstruction, yet perspective observations couple viewpoint utility to camera positions and orientations, increasing the cost of explicit search over candidate poses. Panoramic vision instead provide nearly complete angular coverage and enable position-centric planning. However, exploration in large environments still requires anticipating occluded geometry and allocating computation across candidate spatial trajectories. We present **PIONEER**, a panoramic active mapping framework for efficient large-scale aerial exploration. PIONEER integrates panoramic evidence with a pretrained occupancy prior to estimate cumulative coverage gain over future observation positions. A learned scheduler adapts beam-search width and depth to the mapping state, using predicted runtime and a prescribed planning budget to select search configurations. We also construct a metric-calibrated dataset from 3D assets derived from real-world scenes, providing UE5-compatible textured environments and paired perspective and $360^\circ \times 180^\circ$ panoramic observations for closed-loop evaluation. Experiments demonstrate that PIONEER improves coverage and reconstruction quality under matched exploration time budgets. Dynamic search allocation further reduces the time to reach a common coverage target relative to the evaluated fixed configuration.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Aerial robots provide an efficient platform for exploring and reconstructing large environments, with applications in urban mapping, infrastructure inspection, and disaster response (He et al., 2021; Silano et al., 2021; Tranzatto et al., 2022; Huang et al., 2026). The resulting map quality depends strongly on where observations are collected. Predefined flight paths cannot adapt to scene structures revealed during exploration and may therefore produce redundant observations or incomplete coverage. Active mapping addresses this problem by selecting informative observations online based on the current map (Song et al., 2022; Tao et al., 2023; Feng et al., 2023; Zhang et al., 2025).

However, most existing active mapping methods rely on perspective cameras and formulate viewpoint selection as a joint search over camera position and orientation (Isler et al., 2016; Zhou et al., 2021; Li et al., 2026b). With a limited field of view, the value of an observation depends on both camera position and viewing direction. Explicitly enumerating these choices increases the number of candidate poses evaluated during multi-step trajectory planning.

In contrast, panoramic cameras provide a natural way to reduce this search burden. A $360^\circ \times 180^\circ$ panoramic image provides nearly complete angular coverage at each location (Lin et al., 2025). The predicted observation value therefore depends mainly on the position rather than the camera heading. This allows viewpoint selection to be formulated as a position-centric search. Removing explicit orientation choices makes spatial trajectories the primary search variables. Existing panoramic active mapping methods have demonstrated the potential of this formulation in indoor and small outdoor scenes (Jung et al., 2025; Jin et al., 2026; Chappidi & Zakhor, 2026). Extending this formulation to kilometer-scale aerial environments requires reasoning over a much larger space of candidate positions.

At this scale, full angular coverage does not remove occlusion: the planner must anticipate what future positions may reveal beyond the currently observed surfaces. Evaluating many candidate trajectories also incurs computational cost, while the value of additional search depends on scene structure and exploration progress. Effective exploration therefore calls for both a spatial prior for unseen geometry and adaptive allocation of search effort.

To address these challenges, we present **PIONEER**, a panoramic active mapping framework for efficient aerial exploration in large environments. A pretrained occupancy prior supplies hypotheses about unseen geometry for evaluating cumulative coverage gain along candidate spatial trajectories. Beam width controls the number of retained alternatives, while search depth specifies the number of future translational decisions. A learned scheduler selects both from the mapping state and the available planning budget, using predicted runtime to identify admissible configurations. This makes search effort responsive to the current mapping state and its associated computational cost.

Evaluating this problem requires large environments that support closed-loop observation, planning, and reconstruction at metric scale. Representative panoramic datasets focus on indoor environments (Zheng et al., 2020; Armeni et al., 2017), while large-scale outdoor datasets emphasize reconstruction and novel-view synthesis (Li et al., 2023; Turki et al., 2022). This motivates an evaluation setting that combines panoramic observations with renderable, metric-calibrated scene assets for kilometer-scale aerial active mapping. We therefore construct a dataset using 3D assets derived from real-world scenes. It covers diverse coastal, campus, dense urban, and architectural environments. The dataset provides textured environments compatible with unreal engine, together with paired perspective and $360^\circ \times 180^\circ$ panoramic observations. Experiments show improved coverage and reconstruction quality under matched exploration time budgets, with an additional efficiency benefit from dynamic search allocation.

In summary, our contributions are threefold:

- We formulate panoramic aerial active mapping as position-centric trajectory planning. Full spherical coverage removes explicit orientation enumeration from the ideal coverage objective, while an occupancy prior supports evaluation of future observation positions.
- We develop a learned beam-search scheduler that selects search width and depth from the mapping state and predicted runtime. It improves exploration efficiency over the evaluated fixed configuration by adapting search effort during mapping.

- We construct metric-calibrated outdoor environments with UE5-compatible textured meshes and paired perspective and panoramic observations for closed-loop evaluation. Experiments demonstrate improved coverage and reconstruction quality under matched exploration time budgets.

2 Related Works

Active Mapping. Active mapping selects informative observations to improve scene coverage and reconstruction. Classical methods guide exploration using frontiers or information gain (Bircher et al., 2016; Isler et al., 2016). Hierarchical planners such as FUEL (Zhou et al., 2021) and TARE (Cao et al., 2021) improve scalability by separating local motion planning from global exploration. Recent learning-based methods estimate surface coverage or directly predict future viewpoints. SCONE (Guédon et al., 2022) and MACARONS (Guédon et al., 2023) learn coverage priors, while GenNBV (Chen et al., 2024) learns a generalizable viewpoint policy. NextBestPath (Li et al., 2025a) and MAGICIAN (Li et al., 2026b) further extend the planning horizon by predicting long-term goals or accumulated coverage gain. Nevertheless, most existing methods rely on perspective cameras and search over both camera position and orientation. This enlarges the action space and increases planning cost as the scene scale and search horizon grow. Our work studies position-centric panoramic exploration with adaptive allocation of spatial search effort based on predicted runtime and the mapping state.

Panoramic Vision. Panoramic cameras provide nearly complete angular coverage from each sensing position, reducing the dependence of observation quality on camera heading. A recent survey (Lin et al., 2025) reviews the transition from perspective to panoramic vision and identifies geometric distortion, nonuniform sampling, and boundary continuity as its main challenges. Panoramic observations have been applied to large indoor reconstruction in IM360 (Jung et al., 2025) and autonomous scan planning in Scanning Bot (Lee et al., 2025). R3-RECON (Jin et al., 2026) estimates omnidirectional view utility, while zero-shot active mapping (Wang et al., 2026) combines panoramic observations with bird's-eye-view representations. OPAL (Chappidi & Zakhor, 2026) acquires omnidirectional context through deliberate yaw rotations for aerial exploration, but is evaluated in indoor environments. Recent work has extended panoramic perception to aerial settings. AirSim360 (Ge et al., 2026) provides a panoramic unmanned-aerial-vehicle (UAV) simulator and large-scale aerial observations. Fly360 (Zhang et al., 2026) uses full-view perception for omnidirectional obstacle avoidance. PanoWorld (Li et al., 2026a) further introduces real-world panoramic UAV videos for controllable world generation. These studies motivate panoramic aerial perception; PIONEER focuses on occupancy-guided spatial planning and adaptive search allocation for active mapping in large environments.

Scene Representations for Active Mapping. Active mapping requires a scene representation that supports both online reconstruction and efficient evaluation of unexplored regions. Occupancy anticipation (Ramakrishnan et al., 2020) predicts spatial structure beyond current observations, providing a useful prior for exploration. Neural radiance fields enable continuous reconstruction and uncertainty estimation, as demonstrated by Mega-NeRF (Turki et al., 2022) and NARUTO (Feng et al., 2024), but their online optimization remains expensive. 3D Gaussian Splatting (Kerbl et al., 2023) provides faster rendering and has been adopted by ActiveGS (Jin et al., 2025), ActiveSplat (Li et al., 2025b), ActiveGAMER (Chen et al., 2025), and GAVIS (Xue et al., 2026) for active reconstruction and uncertainty estimation. PanoLOG (Chen et al., 2026) further adapts 3DGS to large-scale outdoor reconstruction from panoramic images through spatial partitioning. However, these representations mainly model observed regions or require repeated scene-specific optimization. They are not designed for fast reasoning over vast unobserved space. PIONEER uses a pretrained occupancy prior and Imagined Gaussians to evaluate candidate spatial trajectories, and learns to select search width and depth according to the mapping state and predicted runtime.

3 Method

The key insight of this paper is that full spherical coverage makes viewing direction redundant in the ideal coverage objective, allowing planning to focus on sequences of observation positions. PI-

ONEER searches over spatial trajectories $\pi_t = (\mathbf{p}_{t+1}, \ldots, \mathbf{p}_{t+H})$, where $\mathbf{p}_{t+h} \in \mathbb{R}^3$ is a candidate observation position and H is the number of future translational decisions considered within one planning call.

3.1 Preliminaries

Active Mapping. Active mapping selects observations sequentially to improve scene coverage. At step t, a UAV acquires an RGB-D observation at camera pose $\mathbf{c}_t \in \mathrm{SE}(3)$. Let $\mathcal{S}_t$ denote the accumulated surface observations and $\mathcal{C}_t = \{\mathbf{c}_1, \ldots, \mathbf{c}_t\}$ the corresponding camera poses. The agent updates its map and plans the next motion based on the available observations.

Occupancy Prediction and Coverage Gain. To estimate the value of unexplored regions, an occupancy model predicts the probability of surface occupancy at a query point $\mathbf{x} \in \mathbb{R}^3$:

$$\widehat{\sigma}_t(\mathbf{x}) = f_{\mathrm{occ}}\left(\mathbf{x}; \mathcal{S}_t, \mathcal{C}_t\right). \tag{1}$$

The predicted field is represented by Imagined Gaussians (Li et al., 2026b). Their opacity encodes occupancy probability, while a novelty attribute indicates whether the associated surface remains unobserved. Rendering these novelty values from a candidate pose provides an efficient estimate of the expected coverage gain.

Beam Search and Trajectory Planning. Planning over multiple future poses accounts for the coverage obtained along a trajectory. Beam search approximates this search by retaining the K partial trajectories with the highest accumulated gain at each of H planning steps, where K and H denote the beam width and planning depth, respectively. Each trajectory maintains an independent novelty state that is updated after every simulated observation to avoid counting the same predicted surface repeatedly. The agent executes the first motion of the selected trajectory and replans after receiving a new observation.

3.2 Panoramic Evidence and Occupancy Prior

Spherical Observation Model. A panoramic observation covers the viewing sphere from a single camera center. To integrate evidence across directions in metric space, its pixels are represented as calibrated rays:

$$\mathcal{R}_t = \left\{(\mathbf{d}^C_{t,i}, \rho_{t,i}, m_{t,i}, h_{t,i}, \omega_{t,i})\right\}_{i=1}^{N_t}, \tag{2}$$

where N_t is the total number of pixels in the panorama at step t, and $i \in \{1, \ldots, N_t\}$ indexes a pixel and its corresponding ray. For pixel i, $\mathbf{d}^C_{t,i}$ denotes the unit ray direction in the camera frame, $\rho_{t,i}$ is the metric distance along the ray, $m_{t,i}$ represents the depth validity indicator, $h_{t,i}$ denotes the surface-hit indicator, and $\omega_{t,i}$ is the solid angle represented by that pixel. Spherical calibration and depth conventions are provided in Appendix A.1.1.

Panoramic Evidence Integration. Spatial reasoning across successive observations requires their geometric evidence to share a common frame. For camera pose $\mathbf{c}_t = (\mathbf{R}_t, \mathbf{p}_t)$, a valid surface hit ($m_{t,i} = h_{t,i} = 1$) gives

$$\mathbf{X}_{t,i} = \mathbf{p}_t + \rho_{t,i}\mathbf{R}_t\mathbf{d}^C_{t,i}, \tag{3}$$

where $\mathbf{X}_{t,i} \in \mathbb{R}^3$ is the observed surface point in world coordinates, and $\mathbf{R}_t$ maps camera directions to the world frame. These registered surface points and the free-space evidence along their rays are fused into the accumulated map:

$$\mathcal{S}_t = \mathrm{Fuse}(\mathcal{S}_{t-1}, \mathcal{R}_t, \mathbf{c}_t). \tag{4}$$

The map $\mathcal{S}_t$ stores surface centroids and observed occupancy masks. Verified rays without a hit contribute free space up to the sensing limit; invalid rays contribute no evidence. Consequently, observations from different directions and positions jointly inform occupancy prediction in a shared spatial map.

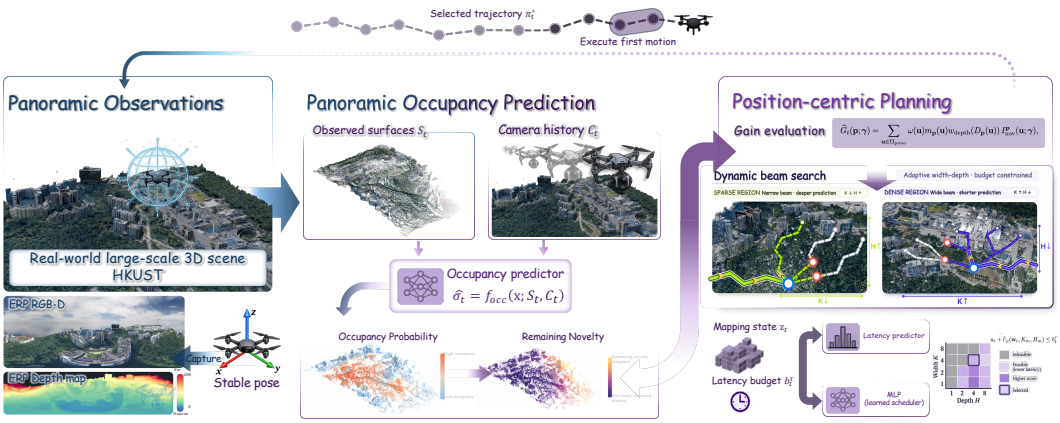

Figure 2: **PIONEER framework.** Panoramic observations update the map and camera history. A pretrained occupancy prior supports trajectory gain evaluation, while a learned scheduler uses the mapping state and predicted runtime to select beam-search width and depth. The first motion is executed before replanning.

Occupancy Prediction. Even with full angular coverage, occluded surfaces remain unobserved. To support coverage evaluation at candidate positions, a pretrained occupancy prior (Li et al., 2026b) infers likely geometry from the accumulated map and camera history $\mathcal{C}_t$, yielding $\widehat{\sigma}_t(\mathbf{x}) = f_{\text{occ}}(\mathbf{x}; \mathcal{S}_t, \mathcal{C}_t)$. Observed occupied and free voxels take precedence over predictions, forming a fused field $\bar{\sigma}_t$. Converting this field into Imagined Gaussians enables repeated coverage evaluation through rasterization. Gaussian opacity encodes occupancy confidence, while novelty identifies unobserved voxels. Known surfaces retain opacity with zero novelty, preserving their occlusion of unseen geometry without adding coverage credit. This representation connects panoramic evidence at visited positions to the potential value of future positions. The prediction interface and Gaussian construction are detailed in Appendix A.1.2.

3.3 Position-Centric Trajectory Planning

Position-Centric Formulation. The full angular coverage of a panorama motivates organizing information search around sensing positions. For fixed geometry and novelty, let $g_t(\mathbf{p}, \mathbf{d})$ denote the directional gain density at position $\mathbf{p}$. A camera with rotation $\mathbf{R}$ and angular support χ has gain

$$G_t(\mathbf{p}, \mathbf{R}) = \int_{\mathbb{S}^2} \chi(\mathbf{R}^{-1}\mathbf{d}) g_t(\mathbf{p}, \mathbf{d})\, \mathrm{d}\omega, \tag{5}$$

where $\mathrm{d}\omega$ denotes the solid angle element on the unit sphere $\mathbb{S}^2$. With complete spherical support and sensing range and validity independent of orientation, $\chi \equiv 1$ gives $G_t(\mathbf{p}, \mathbf{R}) = G_t(\mathbf{p})$. Thus, viewing direction is redundant in the ideal coverage objective. PIONEER exploits this property by evaluating trajectories of observation positions without enumerating viewing directions. A fixed rendering orientation keeps discrete angular sampling consistent. Camera rotation remains part of observation registration, and motions remain subject to feasibility constraints.

Panoramic Coverage Gain. Each candidate position is scored by the novel geometry that a panorama could reveal there. Rendering six cubemap faces of the Imagined Gaussians at $\mathbf{p}$ and resampling them into equirectangular projection (ERP) provides novelty $I^{\mathbf{P}}_{\text{nov}}$, metric range $D_{\mathbf{p}}$, and validity $m_{\mathbf{p}}$ maps. The coverage surrogate is

$$\begin{aligned} \widehat{G}_t(\mathbf{p}; \boldsymbol{\gamma}) &= \sum_{\mathbf{u} \in \Omega_{\text{pano}}} \omega(\mathbf{u}) m_{\mathbf{p}}(\mathbf{u}) w_{\text{depth}}(D_{\mathbf{p}}(\mathbf{u}))\, I^{\mathbf{P}}_{\text{nov}}(\mathbf{u}; \boldsymbol{\gamma}), \\ w_{\text{depth}}(r) &= \min\{1, (r/r_0)^2\}, \end{aligned} \tag{6}$$

where r_0 is a fixed metric reference distance and invalid rays contribute zero. Solid-angle weights compensate for ERP sampling density, while the clipped range weighting reduces the preference for nearby surfaces caused by their large angular footprint. Gaussian opacity and compositing account for occupancy confidence and occlusion, allowing all directions to contribute to one position score.

Beam Search. Revealing occluded geometry may require intermediate moves with little immediate gain. Beam search therefore evaluates the coverage accumulated along a spatial trajectory $\pi_t = (\mathbf{p}_{t+1}, \ldots, \mathbf{p}_{t+H})$:

$$U_t(\pi_t) = \sum_{h=1}^{H} \widehat{G}_t\left(\mathbf{p}_{t+h}; \boldsymbol{\gamma}_t^{(h-1)}\right), \tag{7}$$

where $\boldsymbol{\gamma}_t^{(h-1)}$ is the branch novelty after its first $h-1$ hypothetical observations. Each branch expands through six axis-aligned translations, with infeasible segments removed before scoring. Each child inherits its parent's novelty vector; entries visible from its candidate position are cleared after scoring. This suppresses repeated coverage credit while retaining opacity for occlusion. Geometry is shared across branches; only novelty changes within the search. The K highest scoring paths survive at each depth. After depth H, the UAV executes the first motion of the best path and acquires a new panorama to update the map and predictions before replanning. The executed segment must lie in observed free space; later hypothetical segments may use predicted free space. Search details are given in Appendix A.1.3, and Section 3.4 selects K and H using the planning budget and predicted runtime.

3.4 Dynamic Beam Search Scheduler

Configuration Selection. We adapt the beam width and planning depth to the current mapping state and a prescribed latency budget b_t^τ. A lightweight 3D CNN E_θ encodes a voxel grid of predicted occupancy and observed occupied and free space around the current UAV position into $\mathbf{z}_t$. A separate predictor fitted on the target hardware estimates an upper quantile $\widehat{r}_\psi(\mathbf{w}_t, K, H)$ of the remaining runtime for scheduling, search, and motion output, using observable workload statistics $\mathbf{w}_t$. Let a_t be the elapsed time from receiving the observation to the start of scheduling. For a finite catalogue $\{(K_m, H_m)\}_{m=1}^M$, the admissible configuration indices are

$$\mathcal{F}_t(b_t^\tau) = \{m \in \{1, \ldots, M\} : a_t + \widehat{r}_\psi(\mathbf{w}_t, K_m, H_m) \le b_t^\tau\}. \tag{8}$$

An MLP h_ϕ scores each configuration using $\mathbf{z}_t$ and a normalized version of the remaining budget $b_t^\tau - a_t$. A softmax over $\mathcal{F}_t$ defines probabilities $p_{t,m}$. For a nonempty admissible set, we select

$$m_t = \operatorname*{arg\,max}_{m \in \mathcal{F}_t(b_t^\tau)} p_{t,m}, \qquad (K_t, H_t) = (K_{m_t}, H_{m_t}), \tag{9}$$

run spatial beam search, and execute its first motion before replanning. If the set is empty, the configuration with the shortest predicted runtime is used as a fallback. Admission is based on predicted runtime and does not provide a hard deadline guarantee.

Offline Training. For each training state i and sampled budget b with $\mathcal{F}_i(b) \neq \emptyset$, we compare admissible configurations from the same initial map, UAV pose, and model state. Each configuration is used for the first planning decision only. The remaining $L-1$ motions follow a common reference policy that updates observations and rechecks admission under the same budget b. The utility $J_i(m; b)$ accumulates coverage improvements over the same L executed motions for every configuration, rewarding gains earlier in the motion sequence. This supervision window is independent of search depth H. We form soft targets $q_i(m; b) \propto \exp(J_i(m; b)/T)$ over $\mathcal{F}_i(b)$, where $T > 0$ is a temperature, and jointly train the spatial encoder and selection head:

$$\mathcal{L}_{\text{sched}}(\theta, \phi) = -\mathbb{E}_{(i,b)}\left[\sum_{m \in \mathcal{F}_i(b)} q_i(m; b) \log p_{i,m}(b)\right], \tag{10}$$

where $p_{i,m}(b)$ is the scheduler probability at state i under budget b. The occupancy and latency predictors remain frozen. Ground truth geometry is used only to evaluate offline coverage targets. Latency profiling, network architecture, and rollout supervision are detailed in Appendices A.1.4 and A.1.5.

4 Dataset

To evaluate the proposed method, we construct a dataset of large-scale outdoor scenes for aerial active mapping and 3D reconstruction. The underlying 3D assets are sourced from Open3Dhk[1] and

[1]Hong Kong Lands Department's 3D Digital Map: https://3d.map.gov.hk/.



Figure 3: Representative scenes and aerial panoramic observations from the proposed dataset. Textured meshes and surface-normal visualizations illustrate diverse large-scale outdoor environments, including campuses, urban districts, and natural landscapes.

FUNES[2]. As illustrated in Figure 3, the dataset spans diverse environments, including campuses, urban districts, and natural landscapes. Each scene includes a 3D mesh and its associated textures, providing both geometry and appearance for rendering observations at specified camera poses. We standardize scene geometry and camera poses to a common metric scale, ensuring consistent spatial measurements across scenes.

For each scene, we render $360° \times 180°$ panoramic images and perspective images from multiple aerial viewpoints, and provide the corresponding camera parameters. The perspective observations are accompanied by COLMAP-format files to facilitate reconstruction using existing pipelines (Schönberger & Frahm, 2016; Kerbl et al., 2023; Guédon et al., 2025). Together, the textured scene assets and multi-view observations support a complete workflow from online viewpoint selection and observation acquisition to scene reconstruction and evaluation.

5 Experiments

5.1 Experimental Setup

Settings. We evaluate PIONEER on HKUST, ICC, Ninepin Group, and Kennedy Town. The main comparisons use the same exploration time budget for all methods within each scene. Coverage and reconstruction are evaluated using only observations acquired within that budget. Scene geometry and camera poses are calibrated in meters. All methods use translation increments of 20 m along each horizontal axis and 5 m along the vertical axis. PIONEER uses panoramic RGB-D observations with known camera poses and simulator-provided depth. Each observation comprises six co-located cubemap faces, each rendered at 1024×1024 pixels with a $90° \times 90°$ field of view, jointly providing $360° \times 180°$ angular coverage. Each evaluation run performs scene rendering and online planning on one NVIDIA GeForce RTX 3090 GPU (24 GB). The seven available GPUs are used to parallelize independent runs.

Metrics. We report surface coverage at the common time limit (Cover.) and normalized area under the coverage–time curve (AUC), measuring mapping completeness and coverage accumulation within the same exploration time budget. Reconstruction quality is evaluated using structural similarity (SSIM), peak signal-to-noise ratio (PSNR), learned perceptual image patch similarity (LPIPS), and geometric accuracy (Acc.). The first three measure rendering quality, whereas Acc. measures geometric agreement with the reference mesh. We use MILo (Guédon et al., 2025) to reconstruct meshes from the images acquired within each scene's common exploration time budget.

Baselines. We compare against SCONE (Guédon et al., 2022), which predicts next-view coverage gains through volumetric integration; MACARONS (Guédon et al., 2023), which learns reconstruc-

[2] https://funes.world/.

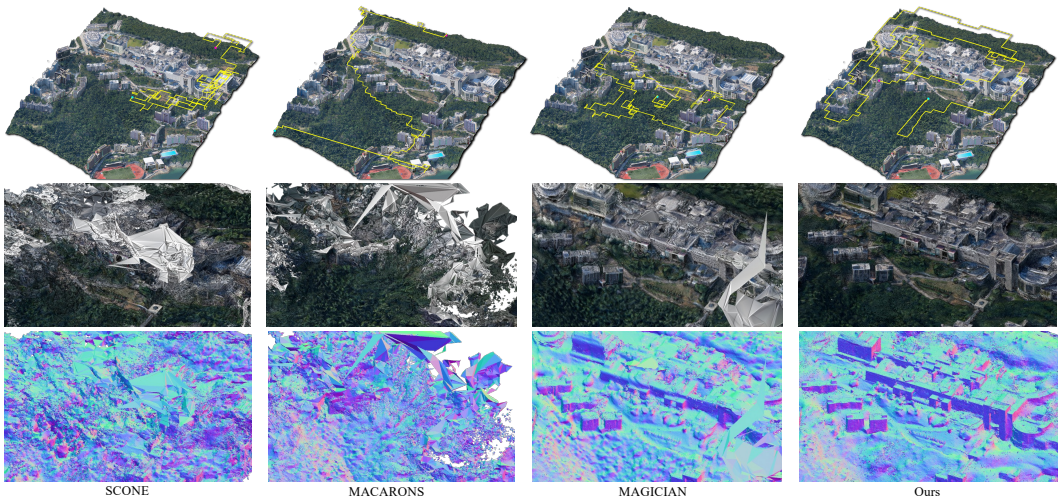

Figure 4: **Qualitative comparison on HKUST.** From top to bottom: exploration trajectories, textured MILo reconstructions, and surface normal maps. PIONEER achieves broader spatial coverage and preserves more complete building geometry with fewer visible artifacts.

tion and coverage anticipation through online RGB self-supervision; and MAGICIAN (Li et al., 2026b), which uses Imagined Gaussians to evaluate trajectory coverage gains in beam search.

5.2 Exploration and Reconstruction Results

Quantitative results. B1 and B2 indicate the top two performed starting points. Under the same exploration time budget within each scene, PIONEER achieves the highest AUC, coverage, PSNR, and geometric accuracy in all eight reported settings in Table 1. Relative to MAGICIAN, coverage increases by 21.03–30.04 percentage points and PSNR by 1.25–4.28 dB. PIONEER also obtains the best SSIM and LPIPS in six of the eight settings for each metric. MAGICIAN retains higher SSIM on Kennedy Town and lower LPIPS at HKUST B1 and Kennedy Town B1. These results show improved coverage and generally better reconstruction quality within the available exploration time. Nevertheless, PIONEER consistently improves exploration performance while delivering broad gains in reconstruction quality.

Table 1: **Exploration and reconstruction under matched time budgets.** All methods share the same exploration time limit within each scene. Dark/light purple denotes the best/second-best result in each scene and reported setting.

Scene	Method	B1						B2					
		AUC↑	Cover. (%)↑	SSIM↑	PSNR↑	LPIPS↓	Acc. (%)↑	AUC↑	Cover. (%)↑	SSIM↑	PSNR↑	LPIPS↓	Acc. (%)↑
HKUST	SCONE	0.278	46.065	0.486	11.737	0.456	91.221	0.233	45.332	0.437	10.826	0.479	89.531
	MACARONS	0.324	48.081	0.478	11.560	0.513	93.902	0.298	50.523	0.482	12.014	0.425	93.272
	MAGICIAN	0.335	70.776	0.545	14.489	**0.369**	98.362	0.332	60.468	0.517	13.962	0.384	96.644
	PIONEER	**0.727**	**91.804**	**0.607**	**17.087**	0.377	**99.417**	**0.681**	**87.656**	**0.579**	**16.257**	**0.381**	**98.460**
ICC	SCONE	0.255	45.263	0.485	10.932	0.490	92.137	0.190	38.381	0.396	10.230	0.494	90.736
	MACARONS	0.134	27.954	0.378	9.171	0.531	75.095	0.111	22.866	0.341	9.058	0.565	72.162
	MAGICIAN	0.325	51.917	0.532	10.724	0.413	98.878	0.331	50.568	0.528	10.836	0.436	97.925
	PIONEER	**0.575**	**79.839**	**0.576**	**15.005**	**0.385**	**98.940**	**0.603**	**76.011**	**0.581**	**14.961**	**0.402**	**98.687**
Ninepin Group	SCONE	0.218	30.367	0.641	11.308	0.440	96.833	0.119	29.742	0.463	10.035	0.558	87.528
	MACARONS	0.150	23.193	0.634	9.750	0.547	89.942	0.094	13.976	0.394	8.609	0.583	87.081
	MAGICIAN	0.319	53.529	0.683	14.056	0.449	97.764	0.307	53.380	0.679	13.924	0.457	97.343
	PIONEER	**0.529**	**83.572**	**0.742**	**16.570**	**0.359**	**98.499**	**0.410**	**82.886**	**0.720**	**16.375**	**0.354**	**98.519**
Kennedy Town	SCONE	0.457	59.086	0.330	10.623	0.543	93.976	0.301	54.885	0.306	10.178	0.592	89.439
	MACARONS	0.157	22.889	0.388	10.792	0.545	72.140	0.148	20.907	0.295	10.720	0.623	70.766
	MAGICIAN	0.409	65.645	**0.424**	12.385	**0.481**	93.326	0.426	58.596	**0.408**	12.197	0.542	93.805
	PIONEER	**0.679**	**89.264**	0.419	**13.781**	0.518	**95.104**	**0.687**	**88.541**	0.397	**13.451**	**0.526**	**94.374**

Qualitative results. Figure 4 compares exploration trajectories and MILo reconstructions on HKUST. PIONEER explores a broader portion of the campus and produces more complete build-

ing structures with fewer prominent geometric artifacts in the displayed region. The normal maps further show better preserved building boundaries and more coherent surfaces. These observations suggest that broader spatial coverage supplies more complete visual evidence for reconstruction. More qualitative experiment results are detailed in Appendix A.2

5.3 Efficiency Study

We compare MAGICIAN, PIONEER with fixed beam width K and search depth H, and PIONEER-Dynamic Beam Search (DBS), which adapts both through the learned scheduler. Figure 5 reports the online time required to reach 50% surface coverage from five starting positions in each scene; NR indicates that the target is not reached. PIONEER-DBS reduces time relative to the fixed variant across settings, demonstrating the efficiency benefit of adapting search effort to the mapping state. Both PIONEER variants also outperform MAGICIAN in the urban scenes whenever MAGICIAN reaches the target, with particularly large gains in Kennedy Town and ICC.

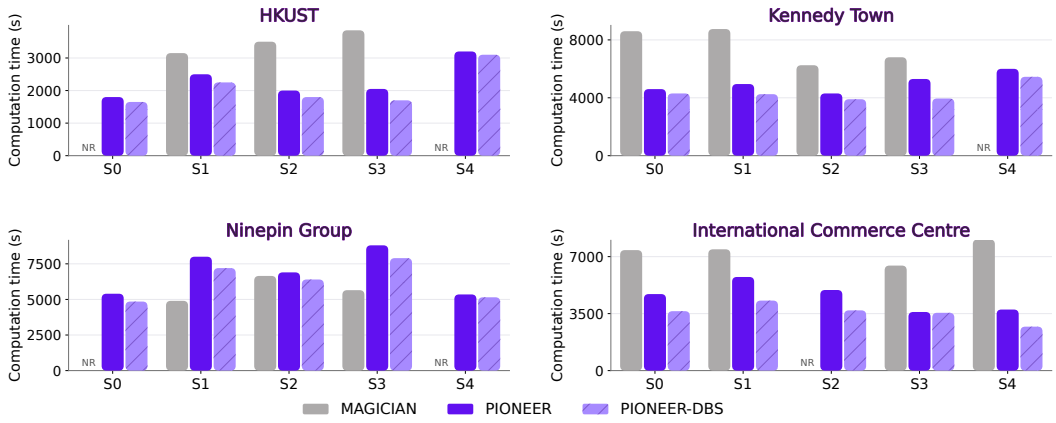

Figure 5: **Exploration efficiency and scheduler ablation.** Time to reach 50% surface coverage from five starting positions per scene. PIONEER (Fix) uses constant beam width and search depth; PIONEER-DBS selects both using the learned scheduler. NR denotes failure to reach the target.

The comparison further reveals a dependence on scene structure. On Ninepin Group, MAGICIAN is faster at S1 and S3, while PIONEER-DBS is slightly faster at S2. The open island layout likely reduces occlusion and makes informative surfaces accessible from fewer viewing directions, limiting the additional benefit of panoramic observations. In urban scenes, surrounding buildings distribute surfaces across directions and conceal others behind intervening structures. Panoramic observations capture multiple surrounding facades, while occupancy-guided spatial search helps select positions that reveal occluded regions. These complementary effects offer a plausible explanation for PIONEER's greater efficiency advantage in urban environments.

6 Conclusion

We presented PIONEER, a panoramic active mapping framework for large-scale aerial exploration. Full angular coverage enables position-centric trajectory planning, while a pretrained occupancy prior supports reasoning about unseen geometry. A learned scheduler uses the mapping state and predicted runtime to select beam-search width and depth. Evaluation in our metric-calibrated environments shows improved coverage and reconstruction quality under matched exploration time budgets, with further efficiency gains from dynamic search allocation. Looking ahead, advances in panoramic depth estimation (Lin et al., 2026) open opportunities to extend PIONEER toward RGB-based panoramic exploration. Integrating these models could reduce reliance on directly measured depth and facilitate deployment with panoramic cameras in real-world aerial environments.

AI USE STATEMENT

In this work, we used generative AI tools only to polish the writing, generate individual graphical elements (e.g., UAV icons) to improve the visual presentation of the framework diagram, and assist with adapting existing code across different scenes. The final teaser figure, framework diagram, and all other figures were manually designed and assembled by the authors, with AI-generated visual content limited to the individual elements described above. We did not use generative AI to conceive the research ideas or develop the initial code framework. We reviewed all AI-assisted work, including textual revisions, graphical elements, and code adaptations, for accuracy and consistency with the intended scientific content and implementation. We take full responsibility for the final content of this work, including all text, claims, code, and visual materials produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

The appendix provides observation geometry, search algorithms, scheduler supervision, evaluation protocols, and implementation settings. These include the time accounting, reconstruction inputs, and data partitions used in our experiments. We will publicly release the source code and dataset after the review decisions are announced.

REFERENCES

Iro Armeni, Sasha Sax, Amir R. Zamir, and Silvio Savarese. Joint 2D-3D-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017. 2

Andreas Bircher, Mina Kamel, Kostas Alexis, Helen Oleynikova, and Roland Siegwart. Receding horizon "next-best-view" planner for 3d exploration. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1462–1468, 2016. 3

Chao Cao, Hongbiao Zhu, Howie Choset, and Ji Zhang. TARE: A hierarchical framework for efficiently exploring complex 3d environments. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi: 10.15607/RSS.2021.XVII.018. URL https://www.roboticsproceedings.org/rss17/p018.html. 3

Yoga Satwik Chappidi and Avideh Zakhor. Opal: Omnidirectional path-efficient aerial 3d exploration, 2026. URL https://arxiv.org/abs/2605.25423. 2, 3

Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. Activegamer: Active gaussian mapping through efficient rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16486–16497, June 2025. 3

Weijian Chen, Weibo Yao, Yuhang Zhang, Xiaolin Tang, Guo Wang, Weijun Zhang, Xitong Gao, Yihao Chen, Hongde Qin, and Lu Qi. Geometry and gradient-based partitioning for panoramic outdoor reconstruction, 2026. URL https://arxiv.org/abs/2607.08769. 3

Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, and Jiangmiao Pang. Gennbv: Generalizable next-best-view policy for active 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16436–16445, June 2024. 3

Chen Feng, Haojia Li, Fei Gao, Boyu Zhou, and Shaojie Shen. PredRecon: A prediction-boosted planning framework for fast and high-quality autonomous aerial reconstruction. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1207–1213, 2023. doi: 10.1109/ICRA48891.2023.10160933. 2

Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, and Yi Xu. Naruto: Neural active reconstruction from uncertain target observations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21572–21583, June 2024. 3

Xian Ge, Yuling Pan, Yuhang Zhang, Xiang Li, Weijun Zhang, Dizhe Zhang, Zhaoliang Wan, Xin Lin, Xiangkai Zhang, Juntao Liang, Xiangtai Li, WenJie Jiang, Bo Du, Ming-Hsuan Yang, and Lu Qi. Airsim360: A panoramic simulation platform within drone view. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26931–26940, June 2026. 3

Antoine Guédon, Pascal Monasse, and Vincent Lepetit. Scone: Surface coverage optimization in unknown environments by volumetric integration. In *Advances in Neural Information Processing Systems*, volume 35, pp. 20731–20743, 2022. doi: 10.52202/068431-1507. 3, 7

Antoine Guédon, Tom Monnier, Pascal Monasse, and Vincent Lepetit. Macarons: Mapping and coverage anticipation with rgb online self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 940–951, June 2023. 3, 7

Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, and Maks Ovsjanikov. MILo: Mesh-in-the-loop gaussian splatting for detailed and efficient surface reconstruction. *ACM Transactions on Graphics*, 2025. URL https://anttwo.github.io/milo/. 7, 19

Jinhao He, Yuming Zhou, Lixiang Huang, Yang Kong, and Hui Cheng. Ground and aerial collaborative mapping in urban environments. *IEEE Robotics and Automation Letters*, 6(1):95–102, 2021. doi: 10.1109/LRA.2020.3032054. 2

Xijie Huang, Weiqi Gai, Tianyue Wu, Congyu Wang, Qiaoyu Zheng, Zhiyang Liu, Xin Zhou, Yuze Wu, and Fei Gao. Navdreamer: Video models as zero-shot 3d navigators. *IEEE Robotics and Automation Letters*, 11(9):10513–10520, 2026. 2

Stefan Isler, Reza Sabzevari, Jeffrey Delmerico, and Davide Scaramuzza. An information gain formulation for active volumetric 3d reconstruction. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3477–3484, 2016. 2, 3

Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activegs: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters*, 10 (5):4866–4873, 2025. 3

Xiaofeng Jin, Matteo Frosi, Yiran Guo, and Matteo Matteucci. R3-recon: Radiance-field-free active reconstruction via renderability, 2026. URL https://arxiv.org/abs/2601.07484. 2, 3

Dongki Jung, Jaehoon Choi, Yonghan Lee, and Dinesh Manocha. Im360: Large-scale indoor mapping with 360 cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 29040–29050, October 2025. 2, 3

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 3, 7

Euijeong Lee, Kyung Min Han, and Young J. Kim. Scanning bot: Efficient scan planning using panoramic cameras. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6403–6409, 2025. 3

Haoyuan Li, Dizhe Zhang, Yuemei Zhou, Xiangkai Zhang, Haoran Feng, Xiaofan Lin, Wenjie Jiang, Bo Du, Ming-Hsuan Yang, and Lu Qi. PanoWorld: Real-world panoramic generation, 2026a. URL https://arxiv.org/abs/2607.09661. 3

Shiyao Li, Antoine Guédon, Clémentin Boittiaux, Shizhe Chen, and Vincent Lepetit. Nextbestpath: Efficient 3d mapping of unseen environments. In *International Conference on Learning Representations (ICLR)*, 2025a. URL https://openreview.net/forum?id=7WaRh4gCXp. 3

Shiyao Li, Antoine Guédon, Shizhe Chen, and Vincent Lepetit. Magician: Efficient long-term planning with imagined gaussians for active mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21606–21615, June 2026b. 2, 3, 4, 5, 8, 13

Yixuan Li, Lihan Jiang, Linning Xu, Yuanbo Xiangli, Zhenzhi Wang, Dahua Lin, and Bo Dai. MatrixCity: A large-scale city dataset for city-scale neural rendering and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3205–3215, 2023. 2

Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. Activesplat: High-fidelity scene reconstruction through active gaussian splatting. *IEEE Robotics and Automation Letters*, 10(8):8099–8106, 2025b. 3

Xin Lin, Xian Ge, Dizhe Zhang, Zhaoliang Wan, Xianshun Wang, Xiangtai Li, Wenjie Jiang, Bo Du, Dacheng Tao, Ming-Hsuan Yang, and Lu Qi. One flight over the gap: A survey from perspective to panoramic vision, 2025. URL https://arxiv.org/abs/2509.04444. 2, 3

Xin Lin, Meixi Song, Dizhe Zhang, Wenxuan Lu, Haodong Li, Bo Du, Ming-Hsuan Yang, Truong Nguyen, and Lu Qi. Depth any panoramas: A foundation model for panoramic depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26835–26844, June 2026. 9

Santhosh K. Ramakrishnan, Ziad Al-Halah, and Kristen Grauman. Occupancy anticipation for efficient exploration and navigation. In *Computer Vision – ECCV 2020*, volume 12350 of *Lecture Notes in Computer Science*, pp. 400–418. Springer, 2020. doi: 10.1007/978-3-030-58558-7_24. URL https://doi.org/10.1007/978-3-030-58558-7_24. 3

Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 7

Giuseppe Silano, Tomas Baca, Robert Penicka, Davide Liuzza, and Martin Saska. Power line inspection tasks with multi-aerial robot systems via signal temporal logic specifications. *IEEE Robotics and Automation Letters*, 6(2):4169–4176, 2021. doi: 10.1109/LRA.2021.3068114. 2

Soohwan Song, Daekyum Kim, and Sunghee Choi. View path planning via online multiview stereo for 3-D modeling of large-scale structures. *IEEE Transactions on Robotics*, 38(1):372–390, 2022. doi: 10.1109/TRO.2021.3083197. 2

Yuezhan Tao, Yuwei Wu, Beiming Li, Fernando Cladera, Alex Zhou, Dinesh Thakur, and Vijay Kumar. SEER: Safe efficient exploration for aerial robots using learning to predict information gain. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1235–1241, 2023. doi: 10.1109/ICRA48891.2023.10160295. 2

Marco Tranzatto, Takahiro Miki, Mihir Dharmadhikari, Lukas Bernreiter, Mihir Kulkarni, Frank Mascarich, Olov Andersson, Shehryar Khattak, Marco Hutter, Roland Siegwart, and Kostas Alexis. Cerberus in the DARPA subterranean challenge. *Science Robotics*, 7(66):eabp9742, 2022. doi: 10.1126/scirobotics.abp9742. 2

Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12922–12931, June 2022. 2, 3

Yuanze Wang, Dianxi Shi, Yuetian Wang, Shiming Song, Haikuo Peng, Chunping Qiu, and Mengzhu Wang. Zero-shot active mapping via fused 360-bev representations and vision–language models. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026. URL https://openreview.net/forum?id=kBZguFRp1B. 3

Shangjie Xue, Jesse Dill, Dhruv Ahuja, Frank Dellaert, Panagiotis Tsiotras, and Danfei Xu. Uncertainty-driven 3d gaussian splatting active mapping via anisotropic visibility field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5014–5026, June 2026. 3

Xiangkai Zhang, Dizhe Zhang, Wenzhuo Cao, Zhaoliang Wan, Yingjie Niu, Lu Qi, Xu Yang, and Zhiyong Liu. Fly360: Omnidirectional obstacle avoidance within drone view, 2026. URL https://arxiv.org/abs/2603.06573. 3

Yichen Zhang, Xinyi Chen, Chen Feng, Boyu Zhou, and Shaojie Shen. FALCON: Fast autonomous aerial exploration using coverage path guidance. *IEEE Transactions on Robotics*, 41:1365–1385, 2025. doi: 10.1109/TRO.2024.3522148. 2

Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. Structured3D: A large photo-realistic dataset for structured 3D modeling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 519–535, 2020. 2

Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2): 779–786, 2021. 2, 3

A Appendix

A.1 Method

This appendix provides the geometric derivations, search analysis, and training protocol underlying PIONEER. We use K for beam width, H for search depth, L for the number of executed motions in an oracle rollout, and b_t^τ for the latency budget of a single planning decision. Observed geometry is updated from acquired measurements. Predicted geometry supports planning, while hypothetical observations modify only the novelty state of the corresponding search branch.

A.1.1 Panoramic Observation Geometry

Spherical projection and integration. We use central equirectangular panoramas with full $360^\circ \times 180^\circ$ coverage. Camera axes point right, down, and forward along x, y, and z, respectively. For a panorama of size $W_{\text{img}} \times H_{\text{img}}$, pixel (u, v) defines

$$\lambda_u = 2\pi \left(\frac{u + 1/2}{W_{\text{img}}} - \frac{1}{2} \right), \qquad \varphi_v = \pi \left(\frac{1}{2} - \frac{v + 1/2}{H_{\text{img}}} \right), \tag{11}$$

and the unit ray

$$\mathbf{d}^C(u, v) = [\cos \varphi_v \sin \lambda_u, \; -\sin \varphi_v, \; \cos \varphi_v \cos \lambda_u]^\top. \tag{12}$$

Horizontal coordinates wrap periodically across the panorama seam. To integrate gain over viewing directions, we assign each pixel its exact solid angle:

$$\omega(u, v) = \frac{4\pi}{W_{\text{img}}} \sin\left(\frac{\pi}{2H_{\text{img}}} \right) \cos \varphi_v, \qquad \sum_{u,v} \omega(u, v) = 4\pi. \tag{13}$$

These weights compensate for the dense angular sampling near the poles, so candidate scores reflect spherical coverage rather than ERP pixel count.

Metric depth and observed map. Our observation interface stores metric ray range $\rho_t(u, v)$ together with validity and surface-hit flags. Poses and depths are exported in a common coordinate convention and in meters. With camera position $\mathbf{p}_t$ and camera-to-world rotation $\mathbf{R}_t$, a valid surface hit gives

$$\mathbf{X}_t(u, v) = \mathbf{p}_t + \rho_t(u, v)\mathbf{R}_t\mathbf{d}^C(u, v). \tag{14}$$

Only finite surface hits in the sensing interval $[r_{\min}, r_{\max}]$ produce surface points. A ray verified by the simulator to have no hit within $r_{\max}$ records free space up to that range without adding a surface. Invalid rays update neither occupancy nor free space, so missing depth is never interpreted as a clear path. We fuse surface points on a fixed metric voxel grid with spacing δ, retaining their centroid in each occupied voxel. Voxels traversed before the hit voxel are marked free; occupied evidence takes precedence over free evidence. The observed representation $\mathcal{S}_t$ stores these surface centroids and occupancy masks. This shared map supplies geometric evidence to occupancy prediction and the observed-space channels to the scheduler.

A.1.2 Occupancy Prediction and Imagined Gaussians

Occupancy prior. Given the accumulated map $\mathcal{S}_t$ and camera history $\mathcal{C}_t$, a pretrained occupancy model (Li et al., 2026b) estimates occupancy confidence at query points, including those in occluded regions:

$$\widehat{\sigma}_t(\mathbf{x}) = f_{\text{occ}}(\mathbf{x}; \mathcal{S}_t, \mathcal{C}_t) \in [0, 1]. \tag{15}$$

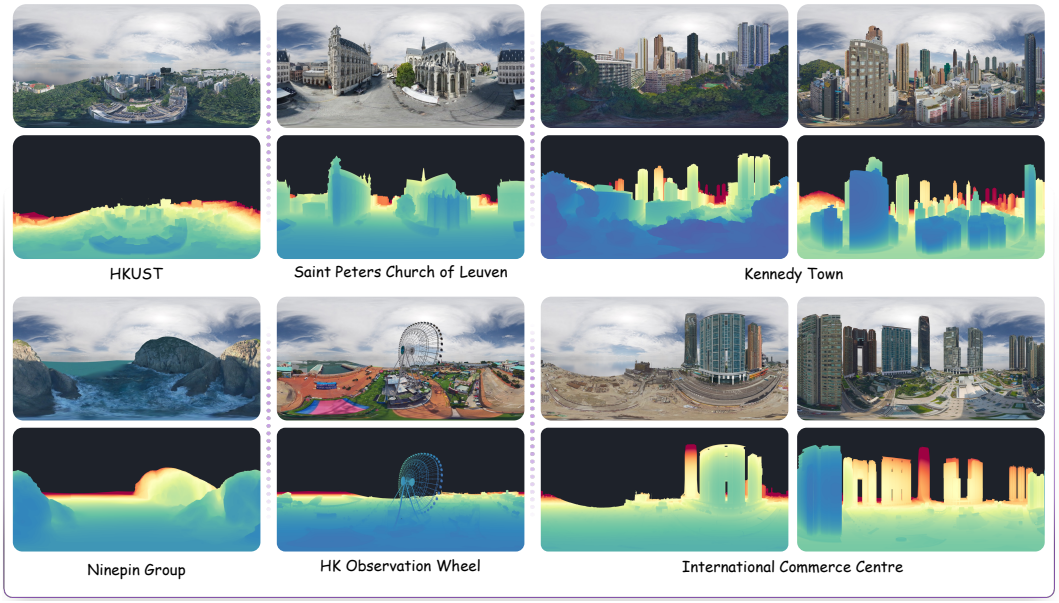

Figure 6: **Panoramic RGB-D observations across diverse scenes.** Each pair shows an RGB panorama (top) and its corresponding color-coded ray-range map (bottom) in equirectangular projection, with $360^\circ \times 180^\circ$ coverage.

We use the pretrained weights directly and keep them fixed throughout scheduler training and exploration. The network consumes surface coordinates and camera history $\mathcal{C}_t$; RGB is reserved for reconstruction. For each query, self-attention and pooling encode its nearest surface neighbors at several point-cloud resolutions. These features are concatenated with a pooled neighborhood feature, a query embedding, and spherical harmonic coefficients of the acquired viewing directions. An MLP produces the occupancy estimate, clipped to $[0, 1]$.

Queries are processed in local prediction boxes. Surface and query coordinates are centered at the box center and divided by its diagonal; view directions use the same coordinate frame. The directional history records acquired rays that reach each query, including measured free space. Empty neighborhoods receive the neutral value $1/2$. This local normalization preserves the pretrained model's input scale while all planning distances remain metric. Predictions are cached by spatial block and refreshed when the block's input surface or viewing history changes; newly activated blocks are queried on demand. The occupancy prior supplies geometric hypotheses; the separately trained 3D CNN scheduler allocates search computation.

A shared Gaussian representation. We query a fixed lattice aligned with the world axes within a local volume covering the maximum supported search horizon and sensing range. For translation length ℓ and maximum depth $H_{\max}$, a volume containing the ball of radius $H_{\max}\ell + r_{\max}$ around $\mathbf{p}_t$ covers all scored rays. Lattice spacing is δ, and observed evidence overrides predictions in the corresponding voxels:

$$\bar{\sigma}_t(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \text{ lies in an observed occupied voxel,} \\ 0, & \mathbf{x} \text{ lies in an observed free voxel,} \\ \widehat{\sigma}_t(\mathbf{x}), & \text{otherwise.} \end{cases} \tag{16}$$

Each retained voxel contains one Gaussian:

$$\mathcal{G}_{t,j} = \left(\mathbf{x}_{t,j}, \frac{\delta^2}{4} \mathbf{I}_3, \bar{\sigma}_t(\mathbf{x}_{t,j}), \gamma_{t,j} \right). \tag{17}$$

Its center is the observed surface centroid when available and the voxel center otherwise. We retain all observed occupied voxels and unobserved voxels whose occupancy exceeds a fixed threshold $\alpha_{\min}$. Novelty is initialized to zero for observed voxels and one for unobserved voxels. Observed

geometry therefore replaces its imagined counterpart instead of introducing a duplicate occluder. Fixed isotropic scales avoid Gaussian optimization during planning and keep the proxy resolution consistent across views.

Panoramic gain rendering. At each candidate position, we rasterize six cubemap faces with a common center and 90° fields of view, then resample them into ERP maps. Each spherical ray selects the face with the largest absolute direction component. Novelty uses bilinear sampling across adjacent faces; depth and validity use nearest sampling to avoid mixing distinct surfaces. With zero background, novelty is composited as

$$I^{\mathbf{p}}_{\mathrm{nov}}(\mathbf{u};\boldsymbol{\gamma}) = \sum_j T^{\mathbf{p}}_j(\mathbf{u}) a^{\mathbf{p}}_j(\mathbf{u}) \gamma_j, \qquad T^{\mathbf{p}}_j(\mathbf{u}) = \prod_{k<j} \left(1 - a^{\mathbf{p}}_k(\mathbf{u})\right), \tag{18}$$

where $a^{\mathbf{p}}_j$ includes Gaussian opacity and its projected footprint, and contributors are ordered from front to back. Compositing stops at $r_{\max}$ along each ray. We use median termination depth, defined by accumulated opacity reaching $1/2$. The depth is invalid if this level is not reached. Face-axis depth z is converted to ray range using $\rho = z/(\mathbf{e}_3^\top \mathbf{d}^F)$, where $\mathbf{d}^F$ is the unit ray in that face's camera frame. This defines the rendered range map $D_{\mathbf{p}}$ and validity mask $m_{\mathbf{p}}$, which also excludes ranges outside $[r_{\min}, r_{\max}]$.

The gain estimator is

$$\begin{aligned} \widehat{G}_t(\mathbf{p};\boldsymbol{\gamma}) &= \sum_{\mathbf{u}} \omega(\mathbf{u}) m_{\mathbf{p}}(\mathbf{u}) w_{\mathrm{depth}}(D_{\mathbf{p}}(\mathbf{u}))\, I^{\mathbf{p}}_{\mathrm{nov}}(\mathbf{u};\boldsymbol{\gamma}), \\ w_{\mathrm{depth}}(r) &= \min\!\left\{1, (r/r_0)^2\right\}, \end{aligned} \tag{19}$$

where r_0 is a fixed metric reference distance. The squared-range factor reduces the preference for nearby surfaces caused by their large angular footprint, while clipping bounds distant contributions. Together with solid-angle weighting, it provides a consistent coverage surrogate for comparing panoramic viewpoints.

Visibility and novelty updates. For Gaussian j, let $r_j = \|\mathbf{x}_j - \mathbf{p}\|_2$ and let $\mathbf{u}_j$ be its spherical projection. We mark it visible when its projected pixel has valid depth, r_j lies in the sensing interval, and $|r_j - D_{\mathbf{p}}(\mathbf{u}_j)| \le \epsilon_d$. We set $\epsilon_d = \sqrt{3}\delta$, the voxel diagonal, to match the visibility tolerance to the spatial resolution of the proxy map. Writing this decision as $v_j(\mathbf{p}) \in \{0,1\}$, each branch updates

$$\gamma'_j = \gamma_j \left(1 - v_j(\mathbf{p})\right) \tag{20}$$

after evaluating its gain. Setting novelty to zero removes subsequent credit while preserving opacity and hence occlusion. Gaussian geometry and opacity are shared and frozen within a search; only the novelty vector is copied for each branch. Hypothetical views never update $\mathcal{S}_t$ or the occupancy model's observation history. After an executed motion, acquired depth updates the map and regenerates the planning representation.

A.1.3 Position-Centric Planning and Search Space Analysis

Why positions suffice. Let $g_t(\mathbf{p}, \mathbf{d})$ be the directional gain density induced by the current geometry and novelty state. For complete spherical support and a rotationally symmetric sensing range, camera rotation only changes the parameterization of the gain integral:

$$G_t(\mathbf{p}, \mathbf{R}) = \int_{\mathbb{S}^2} g_t(\mathbf{p}, \mathbf{R}\mathbf{q})\, \mathrm{d}\omega(\mathbf{q}) = \int_{\mathbb{S}^2} g_t(\mathbf{p}, \mathbf{d})\, \mathrm{d}\omega(\mathbf{d}) = G_t(\mathbf{p}). \tag{21}$$

The equality follows from rotational invariance of spherical measure. It motivates allocating search to spatial positions. All simulated observations and candidate renders use the same canonical orientation, fixing the discrete angular sampling pattern. Camera rotation is still used when registering acquired observations.

Spatial actions and trajectory utility. We use six translations $\ell\{\pm\mathbf{e}_x, \pm\mathbf{e}_y, \pm\mathbf{e}_z\}$ on a metric position lattice. A successor is rejected if the segment leaves the prescribed flight volume or its swept vehicle sphere intersects a voxel with $\bar{\sigma}_t \ge \tau_{\mathrm{occ}}$, where $\tau_{\mathrm{occ}} \in (0,1)$ is a fixed occupancy threshold.

Algorithm 1 Position-Centric Spatial Beam Search

Require: Position $\mathbf{p}_t$, planning Gaussians, novelty $\boldsymbol{\gamma}_t$, width K, depth H
Ensure: First selected motion, or a no-motion status
$\mathcal{B} \leftarrow \{(\mathbf{p}_t, (), 0, \boldsymbol{\gamma}_t)\}$
for $h = 1, \ldots, H$ **do**
  $\mathcal{E} \leftarrow \emptyset$
  for all $(\mathbf{p}, \pi, U, \boldsymbol{\gamma}) \in \mathcal{B}$ **do**
    for all $\mathbf{p}' \in \mathcal{A}_{\text{valid}}(\mathbf{p}, h)$ **do**
      $(\Delta G, \mathbf{v}) \leftarrow \text{Evaluate}(\mathbf{p}', \boldsymbol{\gamma})$
      $\gamma'_j \leftarrow \gamma_j(1 - v_j)$ for every j
      Add $(\mathbf{p}', \pi \parallel \mathbf{p}', U + \Delta G, \boldsymbol{\gamma}')$ to $\mathcal{E}$
    end for
  end for
  if $\mathcal{E} = \emptyset$ **then**
    return no-motion status
  end if
  $\mathcal{B} \leftarrow \text{TopK}(\mathcal{E})$ by accumulated utility
end for
return the first motion of the highest scoring path in $\mathcal{B}$

The executed first segment must additionally lie in observed free space; later hypothetical segments use the frozen occupancy prediction. Thus, long lookahead can consider unobserved regions while the next motion is supported by acquired depth. All feasibility checks use this map, without access to ground truth scene geometry.

Each beam node stores its trajectory, accumulated utility, and novelty vector. For $\pi_t = (\mathbf{p}_{t+1}, \ldots, \mathbf{p}_{t+H})$,

$$U_t(\pi_t) = \sum_{h=1}^{H} \widehat{G}_t\left(\mathbf{p}_{t+h}; \boldsymbol{\gamma}_t^{(h-1)}\right), \tag{22}$$

where each novelty state follows Eq. 20. Paths reaching the same position remain distinct because their previous observations can differ. Revisits are permitted, allowing a trajectory to return through explored space to reach another region. The translation checks are reversible, so a feasible path can always extend by returning along its preceding edge. If the root has no valid successor, the planner returns a no-motion status. Otherwise, beam search retains the K highest scoring paths at each depth and executes the first motion of the best path after depth H. Ties follow a fixed action order.

Search workload. Let $\mathcal{B}_{h-1}$ denote the retained nodes before expansion h and $\mathcal{A}_{\text{valid}}(n)$ their valid successors. Scoring every valid successor before pruning requires

$$N_{\text{eval}}(K, H) = \sum_{h=1}^{H} \sum_{n \in \mathcal{B}_{h-1}} |\mathcal{A}_{\text{valid}}(n)| \le d \sum_{h=0}^{H-1} \min(K, d^h), \tag{23}$$

where $d = 6$ for our spatial actions. The bound becomes exact for a regular tree without feasibility pruning and accounts for initial beam growth. Once the beam is filled, each additional depth requires at most dK candidate evaluations. For a matched catalogue pairing each spatial successor with A_{ori} orientations, the branching factors are $d_{\text{persp}} = 6A_{\text{ori}}$ and $d_{\text{pano}} = 6$. Removing orientation enumeration therefore makes more spatial paths or greater depth affordable in terms of candidate evaluations. Since panoramic rendering has its own cost, actual width and depth are chosen using the measured latency model in Appendix A.1.4.

In Algorithm 1, $\mathcal{A}_{\text{valid}}(\mathbf{p}, h)$ applies the additional observed free-space requirement when $h = 1$. Evaluate returns gain and visibility before any novelty update; $\parallel$ appends a position. TopK retains all entries when fewer than K are available.

A.1.4 Planning Latency Measurement and Prediction

Latency accounting. We measure planning latency from receiving a panoramic observation to producing the next motion command. At the start of scheduling, map integration, occupancy inference, Imagined Gaussian construction, and scheduler input preparation have already completed.

Their elapsed time a_t is measured directly. We therefore predict only the remaining runtime:

$$\widehat{\tau}_{t,m}^{\mathrm{plan}} = a_t + \widehat{r}_\psi(\mathbf{w}_t, K_m, H_m), \tag{24}$$

where $\{(K_m, H_m)\}_{m=1}^{M}$ is the catalogue of supported search configurations. The remaining runtime includes spatial encoding, latency prediction, configuration selection, beam search, and command preparation. This decomposition uses the time already spent to determine how much computation remains affordable. Physical flight and image capture time are outside this interval.

Latency predictor. The descriptor $\mathbf{w}_t$ contains the numbers of observed and imagined Gaussians used for planning, mean predicted occupancy, and the observed voxel fraction in the scheduler grid. These statistics are available before search. A separate MLP takes normalized $\mathbf{w}_t$, K, and H as inputs and uses two hidden layers with ReLU activations and a softplus output to predict a nonnegative runtime. Keeping K and H as separate inputs captures differences between wide and deep searches with the same product KH.

We fit the predictor using repeated measurements from restored mapping states after warm-up. Profiling executes the full scheduling computation and forces each catalogue configuration in turn. Wall-clock measurements synchronize GPU work at the timing boundaries. To account for runtime variation, we predict an upper conditional quantile using the pinball loss:

$$\begin{aligned} \mathcal{L}_{\mathrm{lat}}(\psi) &= \mathbb{E}_{(\mathbf{w},K,H,r)\sim\mathcal{D}_{\mathrm{lat}}}\left[\rho_\alpha(r - \widehat{r}_\psi(\mathbf{w}, K, H))\right], \\ \rho_\alpha(u) &= \max\{\alpha u, (\alpha - 1)u\}, \end{aligned} \tag{25}$$

where r is the measured remaining runtime and $\alpha \in (1/2, 1)$ specifies the chosen upper quantile. Feature normalization is fixed using the profiling data. The predictor shares no trainable parameters with the scheduler.

Budget admission. For a prescribed planning budget b_t^τ, the admissible configuration indices are

$$\mathcal{F}_t(b_t^\tau) = \{m \in \{1, \dots, M\} : a_t + \widehat{r}_\psi(\mathbf{w}_t, K_m, H_m) \le b_t^\tau\}. \tag{26}$$

The scheduler selects only from this set. If it is empty, we execute the configuration with the smallest predicted runtime and record the decision as having no admissible configuration. Quantile prediction provides an empirical allowance for runtime variation; actual budget violations are measured separately and no hard deadline guarantee is assumed.

A.1.5 Dynamic Beam Search Scheduler and Offline Training

Spatial encoding and configuration selection. The scheduler learns how local geometry and exploration progress affect the value of search width and depth. From predicted occupancy $\widehat{\sigma}_t$ and registered observations $\mathcal{S}_t$, we construct a coarse voxel grid $\mathbf{X}_t$ centered at the current position $\mathbf{p}_t$ and aligned with the world axes. Its fixed metric extent covers the maximum supported planning range and the associated observation region. Three channels encode predicted occupancy, observed occupied space, and observed free space. The latter two are accumulated from depth rays, distinguishing unexplored regions from confirmed empty space.

A spatial encoder E_θ uses three 3D convolution blocks with ReLU activations and spatial downsampling. Adaptive average pooling to a coarse spatial grid followed by flattening retains the arrangement of nearby structures. An MLP h_ϕ with two hidden ReLU layers combines this representation with the remaining budget and outputs one logit per configuration:

$$\begin{aligned} \mathbf{X}_t &= \mathrm{Voxelize}(\widehat{\sigma}_t, \mathcal{S}_t, \mathbf{p}_t), \qquad \mathbf{z}_t = E_\theta(\mathbf{X}_t), \\ \beta_t &= \frac{b_t^\tau - a_t}{b_{\mathrm{ref}}}, \qquad \mathbf{s}_t = h_\phi([\mathbf{z}_t; \beta_t]) \in \mathbb{R}^M, \end{aligned} \tag{27}$$

where $b_{\mathrm{ref}} > 0$ is a fixed time scale. For a nonempty $\mathcal{F}_t = \mathcal{F}_t(b_t^\tau)$, the masked distribution is

$$p_{t,m} = \frac{\exp(s_{t,m})}{\sum_{j\in\mathcal{F}_t}\exp(s_{t,j})}, \qquad m \in \mathcal{F}_t, \tag{28}$$

with zero probability outside $\mathcal{F}_t$. Online execution selects $m_t = \arg\max_{m\in\mathcal{F}_t} p_{t,m}$, runs beam search once with (K_{m_t}, H_{m_t}), and executes its first motion. The configuration is selected again after the next observation. A fixed catalogue order resolves ties throughout.

Algorithm 2 Offline Training of the Dynamic Beam Search Scheduler

Require: Mapping states, configuration catalogue, fitted latency predictor, budgets, window L, temperature T
Ensure: Encoder and scheduler parameters (θ, ϕ)

```
Initialize training set D ← ∅
for all state and budget pairs (i, b) do
    Form F_i(b) using the recorded preprocessing time a_i
    if F_i(b) ≠ ∅ then
        for all m ∈ F_i(b) do
            Restore state i and execute one motion using configuration m
            Execute L − 1 further motions using the reference policy at budget b
            Compute J_i(m; b) using Eq. 29
        end for
        Construct q_i(·; b) and store the training example in D
    end if
end for
Optimize (θ, φ) using Eq. 31
```

Rollout supervision. We collect mapping states from training scenes using several fixed search configurations at different exploration stages. For each state i and sampled budget b, every admissible configuration is evaluated from the same pose, observed map, and model state. Configuration m is used for the first planning decision, and only its first selected motion is executed. The remaining $L-1$ motions follow a common reference policy under the same budget b. At each subsequent decision, this policy updates the map, recomputes admission, and selects the admissible configuration with the largest predicted runtime, using the same fallback when necessary. This rule uses the available planning budget and provides a consistent continuation for evaluating the initial choice.

Let $C_i \in [0,1]$ be initial surface coverage and $C_{i,\ell}^{m,b}$ the coverage after ℓ executed motions under this protocol. We assign the utility

$$J_i(m;b) = \sum_{\ell=1}^{L} \left[C_{i,\ell}^{m,b} - C_i \right]. \tag{29}$$

The common window L is independent of search depth H, so deeper searches receive no advantage from being evaluated over more motions. Accumulating coverage improvements also rewards gains obtained earlier in the window. Ground truth surfaces are used only to measure coverage, never as scheduler inputs or for candidate scoring. The resulting labels evaluate the current configuration under a shared continuation policy; they depend on b because later admission decisions also use this budget.

Training objective. For $\mathcal{F}_i(b) \neq \emptyset$, rollout utilities define soft targets over admissible configurations:

$$q_i(m;b) = \frac{\exp(J_i(m;b)/T)}{\sum_{j \in \mathcal{F}_i(b)} \exp(J_i(j;b)/T)}, \qquad m \in \mathcal{F}_i(b), \tag{30}$$

where $T > 0$ controls target concentration. Soft targets retain alternatives with similar coverage utility. The encoder and selection head are trained jointly with masked cross entropy:

$$\mathcal{L}_{\text{sched}}(\theta, \phi) = -\mathbb{E}_{(i,b)\sim\mathcal{D}} \left[\sum_{m \in \mathcal{F}_i(b)} q_i(m;b) \log p_{i,m}(b) \right], \tag{31}$$

where $p_{i,m}(b)$ is the scheduler probability at state i under budget b. The occupancy predictor and fitted latency predictor remain frozen. The loss learns the relative coverage utility of configurations admitted by the runtime model. Runtime prediction determines admission, while rollout supervision evaluates coverage over a common motion window. Exploration time is evaluated separately in the efficiency study.

A.2 Experiment

We provide additional qualitative comparisons with SCONE, MACARONS, and MAGICIAN on HKUST, ICC, Kennedy Town, and Ninepin Group. These comparisons follow the same scene-

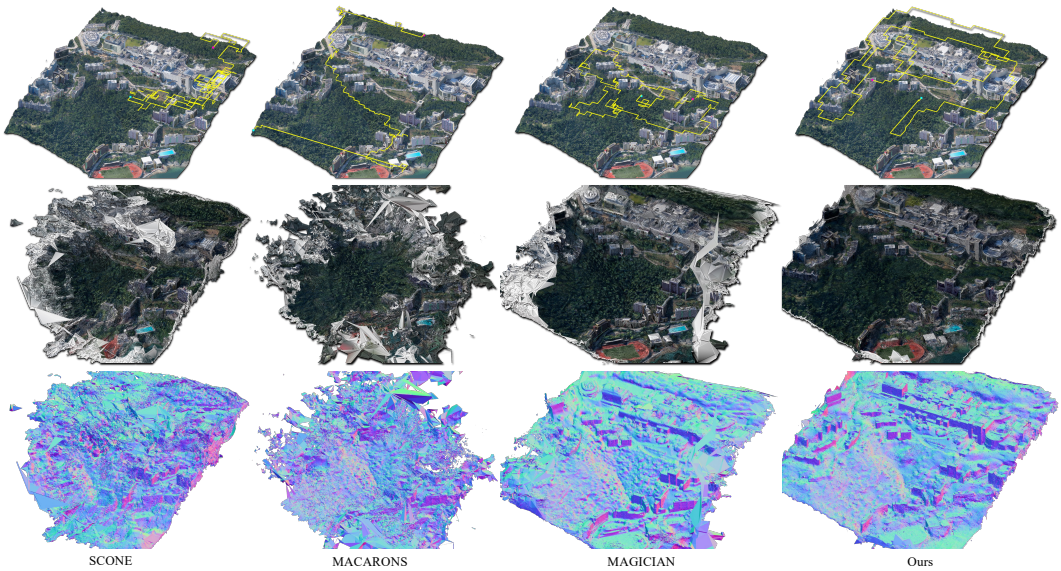

Figure 7: **Qualitative comparison on HKUST.** From top to bottom: exploration trajectories, textured MILo reconstructions, and surface normal maps. PIONEER achieves broader spatial coverage and preserves more complete building geometry with fewer visible artifacts.

specific exploration time budgets as the main experiments. For each method, MILo (Guédon et al., 2025) reconstructs the scene using only images acquired within the corresponding time budget. HKUST, Kennedy Town, and ICC each span $750\,\mathrm{m} \times 600\,\mathrm{m}$, while Ninepin Group spans $1200\,\mathrm{m} \times 1000\,\mathrm{m}$. Figures 7, 8, 9, and 10 show the exploration trajectories, textured reconstructions, and corresponding surface-normal renderings, illustrating how the observations collected within the available time affect reconstruction completeness and structural quality.

Motivation and system design. PIONEER addresses efficient active mapping of large-scale environments under limited exploration time. Our approach jointly considers how observations are acquired and how future observation locations are selected. Panoramic sensing captures all viewing directions at each location, reducing the need to enumerate camera orientations and allowing planning to focus on spatial viewpoint sequences. PIONEER combines this observation model with occupancy-guided multi-step planning and budget-aware allocation of search computation. The broader angular coverage of panoramic observations is an intended advantage of this design and forms part of our contribution. Our experiments therefore assess how these components work together to improve mapping efficiency in large-scale scenes.

Evaluation rationale. The main comparisons evaluate how much of a scene each system can map, and how well it can reconstruct that scene, within the same available time. We use a common time budget for all methods within each scene, with the same per-run computing resources. Each method retains its sensing and planning strategy, and coverage and reconstruction are evaluated using only observations acquired within the budget. This provides a common resource constraint for comparing complete mapping systems while preserving their respective design advantages. Differences in the number and angular coverage of acquired images reflect how each system uses the available time to gather information; the resulting benefits are included in the evaluation.

Time budgets. For the comparisons under fixed time budgets, we allocate total computation times of $5{,}000\,\mathrm{s}$, $10{,}000\,\mathrm{s}$, $12{,}000\,\mathrm{s}$, and $12{,}000\,\mathrm{s}$ to HKUST, ICC, Ninepin Group, and Kennedy Town, respectively. All methods and starting positions share the same budget within each scene. These budgets provide additional exploration time beyond the recorded times to reach 50% coverage in Figure 5. Coverage and reconstruction are evaluated using only observations acquired within the corresponding budget. The supplementary coverage-progression analysis uses a fixed budget of 300 observation locations. This analysis characterizes how coverage accumulates with successive

observations and whether its growth slows as exploration proceeds. Together, the two protocols assess mapping quality within a common time budget and coverage development over a fixed number of observation locations.

Reconstruction settings. The original scene meshes contain approximately 10^7 faces, together with high-resolution textures. Our experiments are conducted on an NVIDIA RTX 3090 GPU with 24 GB of memory. To accommodate MILo reconstruction within this memory budget, we simplify the meshes to approximately 10^6 faces and reduce the texture-atlas configuration from 32×32 to 8×8 for the reconstruction results presented here. These reduced settings limit visual sharpness and the recovery of fine geometric and texture details. The resulting reconstructions serve as a complementary qualitative assessment of exploration coverage, with our comparison focusing on scene completeness and the recovery of major structures. The dataset retains the original full-resolution meshes and textures, supporting reconstruction evaluation at higher fidelity when sufficient GPU memory is available.

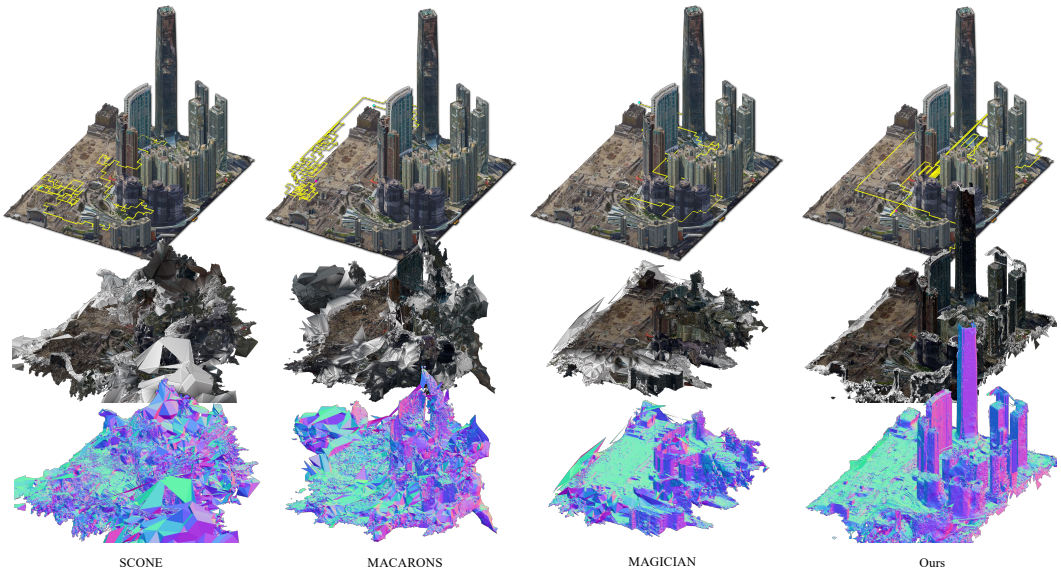

Figure 8: **Qualitative comparison on ICC.** From top to bottom: exploration trajectories, textured MILo reconstructions, and surface normal maps. PIONEER reconstructs the high-rise buildings more completely, particularly their vertical structures.

It can be observed that, PIONEER produces more complete and coherent reconstructions across all four scenes. On HKUST, it better preserves campus buildings and surrounding terrain, with fewer large spurious mesh structures. On ICC, it recovers the dominant high-rise buildings and their vertical extent more completely, whereas the baselines exhibit missing upper structures and substantial geometric artifacts. On Ninepin Island, PIONEER reconstructs the major islands and their coastlines more completely, while the baseline reconstructions contain large missing regions or fragmented surfaces. The surface-normal renderings further illustrate these differences in geometric completeness and structural coherence. These improvements are consistent with the combination of panoramic angular coverage and occupancy-guided long-horizon planning: the former captures observations across viewing directions at each position, while the latter guides exploration toward informative regions beyond the current observations. Together, they enable more effective data acquisition for downstream reconstruction under the same number of exploration steps and total travel distance.

Exploration trajectories and scene-dependent coverage. Figure 11 presents PIONEER's exploration trajectories in four scenes, panoramic observations captured at steps 1, 45, 75, 150, 225, and 300, and the accumulated point clouds after 300 exploration steps. These examples span campus buildings and hillside terrain at HKUST, dense urban structures in Kennedy Town, high-rise buildings and open surrounding areas at the ICC, and island terrain in the Ninepin Group. The ob-

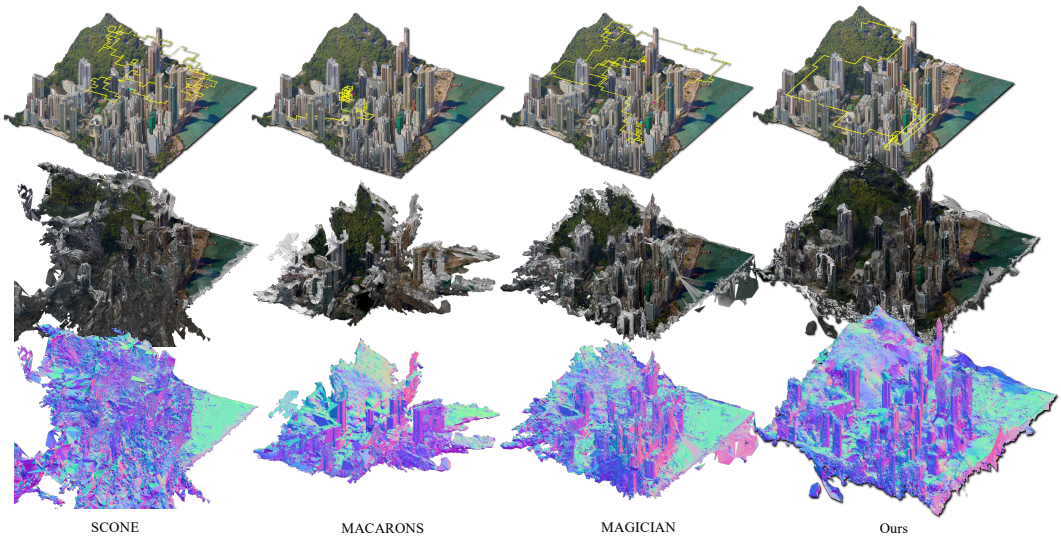

Figure 9: **Qualitative comparison on Kennedy Town.** From top to bottom: exploration trajectories, textured MILo reconstructions, and surface normal maps. PIONEER reconstructs the dense urban buildings more completely.

servations illustrate how translation provides complementary views of scene surfaces despite the full angular coverage available at each position.

After 300 observation locations, the remaining unobserved regions in ICC are consistent with mutual occlusion among closely spaced high-rise buildings. Foreground structures restrict visibility of recessed facades, lower building surfaces, and spaces between buildings, requiring observations from additional positions. Panoramic sensing captures all viewing directions at each position, but cannot reveal surfaces behind intervening structures. Successive observations may therefore overlap substantially while adding relatively little new coverage. From an occupancy perspective, the spatial arrangement of occupied regions determines both the potential value of a candidate position and the visibility of surrounding surfaces. Occupancy predictions guide exploration toward potentially informative regions, while only directly observed surfaces contribute to measured coverage.

The more open areas in HKUST and the exposed terrain in the Ninepin Group offer broader lines of sight, allowing observations to capture more spatially distributed surfaces. Terrain self-occlusion nevertheless requires complementary viewpoints around hills and cliffs. These qualitative results illustrate how scene geometry, visibility, and viewpoint accessibility shape coverage accumulation over 300 observation locations.

Exploration Trajectories. Figure 12 compares the horizontal projections of exploration trajectories from four initial positions (S0–S3) in each scene. PIONEER generally traverses a broader spatial extent, whereas SCONE and MACARONS often produce more concentrated projected trajectories. This difference is particularly visible in Ninepin Group, where PIONEER connects widely separated regions while the trajectories of these two baselines remain largely localized. MAGICIAN also exhibits substantial spatial exploration in several runs, although its trajectories vary considerably with the initial position. Together with the coverage results, these trajectories show that PIONEER acquires observations across more spatially distributed regions in the displayed runs. Together with the coverage results below, these trajectories suggest that the search effectively directs observations toward additional scene regions. Since altitude is omitted, intersections in these projections do not necessarily indicate repeated visits to the same 3D position.

Coverage Progression. Figure 13 plots coverage against the number of observation locations for runs extending to 300 locations. PIONEER maintains higher coverage over most of the plotted range. Coverage growth slows in the later stages on HKUST, Kennedy Town, and ICC, whereas

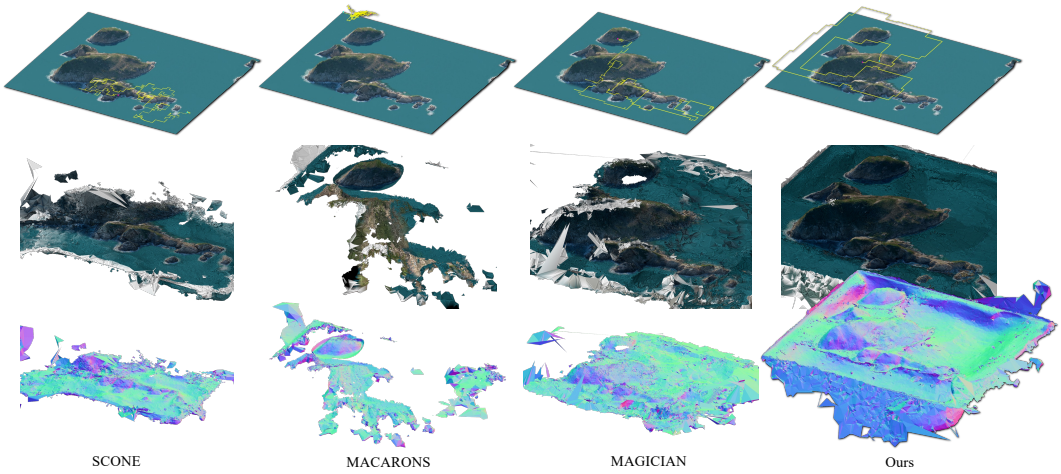

Figure 10: **Qualitative comparison on Ninepin Group.** From top to bottom: exploration trajectories, textured MILo reconstructions, and surface normal maps. PIONEER achieves more complete reconstruction of the major islands and coastlines, with fewer missing regions.

Ninepin Group continues to gain coverage near the end of the evaluated range. These results characterize coverage accumulation and saturation trends under a fixed observation-location budget.

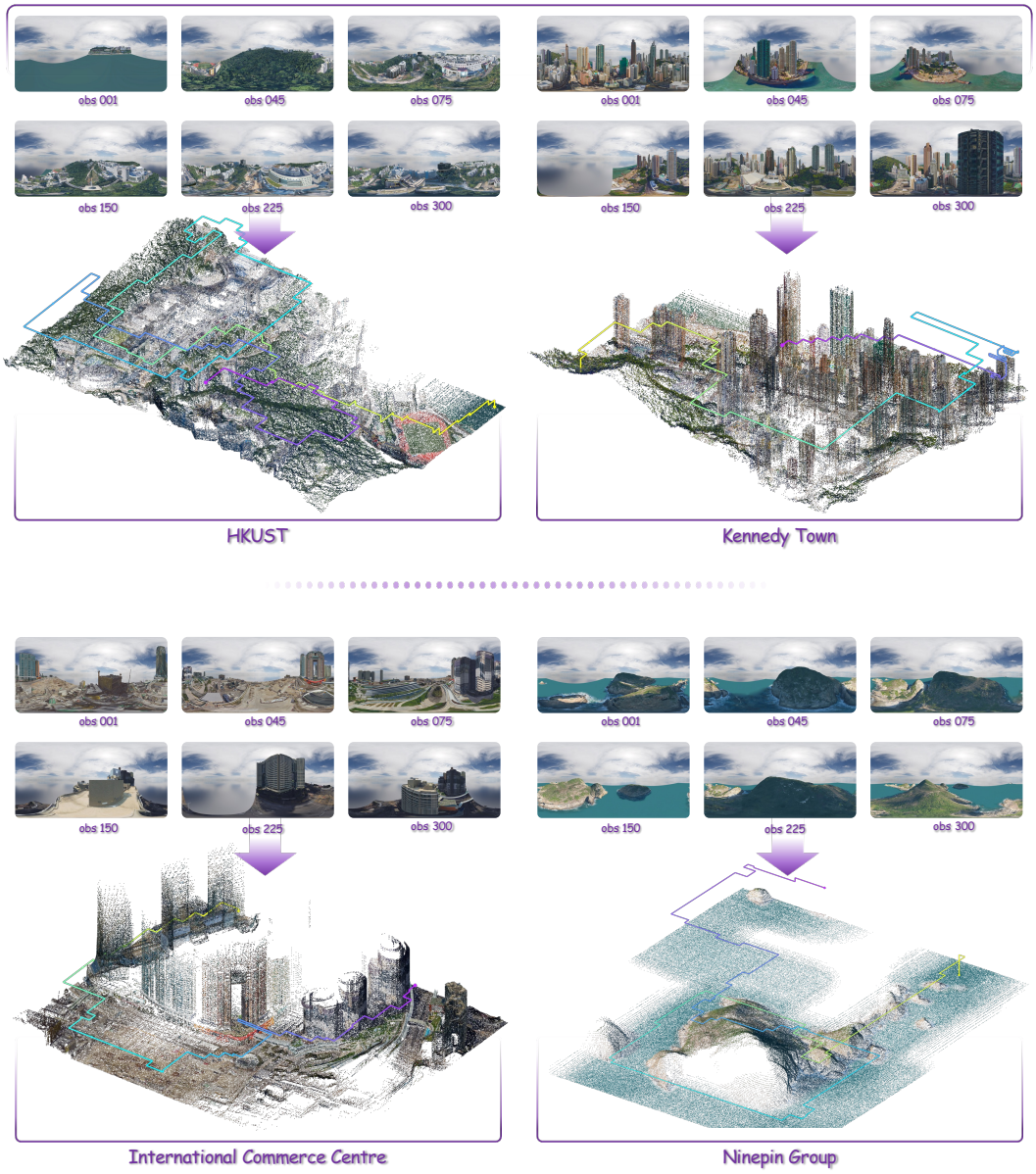

Figure 11: **Qualitative exploration results across four scenes.** For each scene, we show PIONEER's executed trajectory, panoramic observations captured at steps 1, 45, 75, 150, 225, and 300, and the accumulated point cloud after 300 exploration steps.

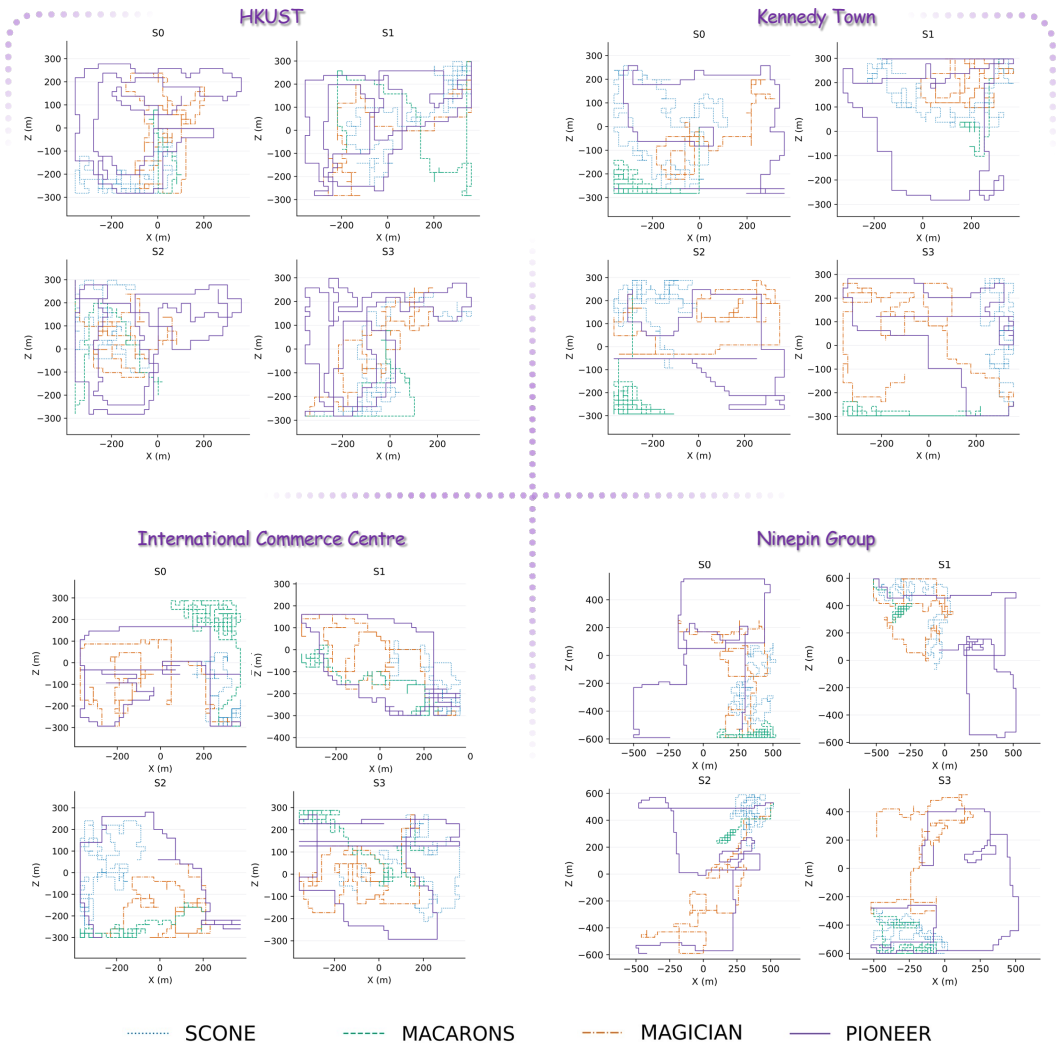

Figure 12: **Exploration trajectories across four scenes.** Horizontal projections are shown for four initial positions (S0–S3) per scene. Colors and line styles distinguish methods; both axes are expressed in meters.

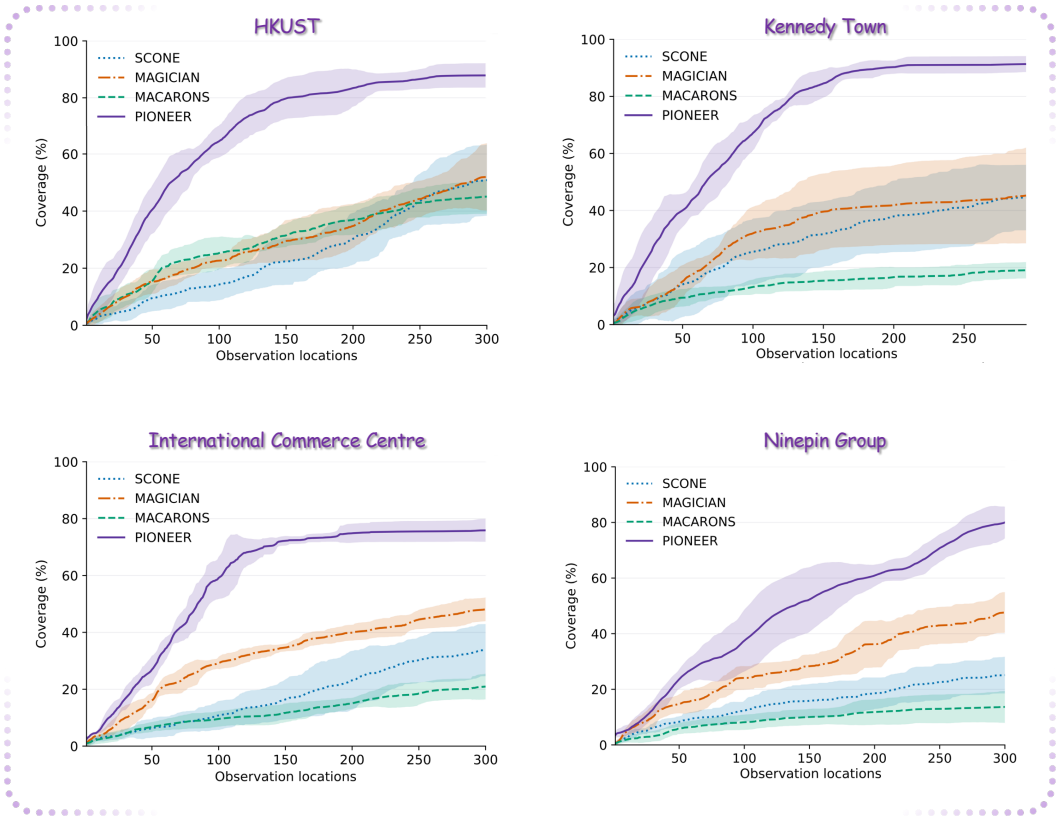

Figure 13: **Surface coverage during exploration.** Coverage is plotted against the number of observation locations for PIONEER and three baselines across four scenes.